Comprehensive transcriptomic and metabolomic analysis of autumn leaf color change in *Fraxinus angustifolia*

Yanlong Wang [Equal first author, 1], **Jinpeng Zhen** [Equal first author, 2], **Xiaoyu Che** [1], **Kang Zhang** [2], **Guowei Zhang** [3], **Huijuan Yang** [1], **Jing Wen** [1], **Jinxin Wang** [1], **Jiming Wang** [1, 4], **Bo He** [5], **Ailong Yu** [6], **Yanhui Li** [Corresp., 1], **Zhigang Wang** [Corresp. 1]

[1] College of Landscape Architecture and Tourism, Hebei Agricultural University, Baoding, China

[2] Hebei Key Laboratory of Plant Physiology and Molecular Pathology, Hebei Bioinformatics Utilization and Technological Innovation Center for Agricultural Microbes, Hebei Agricultural University, Baoding, China

[3] Hongyashan State-owned Forest Farm in Hebei province, Baoding, China

[4] College of Grammar, Hebei Normal University of Science and Technology, Qinhuangdao, China

[5] Green Building Development Center of Baoding, Baoding, China

[6] Flower and wood Technical Service Center of Hengshui, Hengshui, China

Corresponding Authors: Yanhui Li, Zhigang Wang
Email address: yanhuili01@163.com, wzhg1956@163.com

Fraxinus angustifolia is a type of street tree and shade tree with ornamental value. It has a beautiful shape and yellow or reddish purple autumn leaves, but its leaf color formation mechanism and molecular regulation network need to be studied. In this study, we integrated the metabolomes and transcriptomes of stage 1 (green leaf) and stage 2 (red-purple leaf) leaves at two different developmental stages to screen differential candidate genes and metabolites related to leaf color variation. The results of stage 1 and stage 2 transcriptome analysis showed that a total of 5827 genes were differentially expressed, including 2249 upregulated genes and 3578 downregulated genes. Through functional enrichment analysis of differentially expressed genes, we found that they were involved in flavonoid biosynthesis, phenylpropanoid biosynthesis, pigment metabolism, carotene metabolism, terpenoid biosynthesis, secondary metabolite biosynthesis, pigment accumulation, and other biological processes. By measuring the metabolites of *Fraxinus angustifolia* leaves, we found the metabolites closely related to the differentially expressed genes in two different periods of *Fraxinus angustifolia*, among which flavonoid compounds were the main differential metabolites. Through transcriptome and metabolomics data association analysis, we screened nine differentially expressed genes related to anthocyanins. Transcriptome and qRT-PCR results showed that these nine genes showed significant expression differences in different stages of the sample, and we speculate that they are likely to be the main regulatory factors in the molecular mechanism of leaf coloration. This is the first time that we have analyzed the transcriptome combination metabolome in the process of leaf coloration of *Fraxinus angustifolia*, which has important

guiding significance for directional breeding of colored-leaf *Fraxinus* species and will also give new insights for enriching the landscape.

Comprehensive transcriptomic and metabolomic analysis of autumn leaf color change in *Fraxinus angustifolia*

Yanlong Wang[1,#], Jinpeng Zhen[2,#], Xiaoyu Che[1], Kang Zhang[2], Guowei Zhang[3], Huijuan Yang[1], Jing Wen[1], Jinxin Wang[1], Jiming Wang[1,6], Bo He[4], Ailong Yu[5], Yanhui Li[1,*], Zhigang Wang[1,*]

1 College of Landscape Architecture and Tourism, Hebei Agricultural University, Baoding, 071000, China

2 Hebei Key Laboratory of Plant Physiology and Molecular Pathology, Hebei Bioinformatics Utilization and Technological Innovation Center for Agricultural Microbes, Hebei Agricultural University, Baoding, 071000, China

3 Hongyashan State-owned Forest Farm in Hebei province, Baoding, 071000, China

4 Green Building Development Center of Baoding, Baoding, 071000, China

5 Flower and wood Technical Service Center of Hengshui, Hengshui, 053000, China

6 College of Grammar, Hebei Normal University of Science and Technology, Qinhuangdao, 066000, China

# These authors contributed equally to this work

* Corresponding authors:

Yanhui Li

E-mail: yanhuili01@163.com

Zhigang Wang

E-mail: wzhg1956@163.com

Abstract

Fraxinus angustifolia is a type of street tree and shade tree with ornamental value. It has a beautiful shape and yellow or reddish-purple autumn leaves, but its leaf color formation mechanism and molecular regulation network need to be studied. In this study, we integrated the metabolomes and transcriptomes of stage 1 (green leaf) and stage 2 (red-purple leaf) leaves at two different developmental stages to screen differential candidate genes and metabolites related to leaf color variation. The results of stage 1 and stage 2 transcriptome analysis showed that a total of 5827 genes were differentially expressed, including 2249 upregulated genes and 3578 downregulated genes. Through functional enrichment analysis of differentially expressed genes, we found that they were involved in flavonoid biosynthesis, phenylpropanoid biosynthesis, pigment metabolism, carotene metabolism, terpenoid biosynthesis, secondary metabolite biosynthesis, pigment accumulation, and other biological processes. By measuring the metabolites of *Fraxinus angustifolia* leaves, we found the metabolites closely related to the differentially expressed genes in two different periods of *Fraxinus angustifolia*, among which flavonoid compounds were the main differential metabolites. Through transcriptome and metabolomics data association analysis, we screened nine differentially expressed genes related to anthocyanins. Transcriptome and qRT-PCR results showed that these nine genes showed significant expression differences in different stages of the sample, and we speculate that they are likely to be the main regulatory factors in the molecular mechanism of leaf coloration. This is the first time that we have analyzed the transcriptome combination metabolome in the process of leaf coloration of *Fraxinus angustifolia*, which has important guiding significance for directional breeding of colored-leaf *Fraxinus* species and will also give new insights for enriching the landscape.

Keywords: *Fraxinus angustifolia*; transcriptomes; metabolomes; anthocyanin; gene expression

1. Introduction

Colored-leaf plants refer to plants whose leaves are stable, non-green, and of obvious ornamental value in certain seasons or throughout the year (Wang et al. 2016). As important as flower color, leaf color is a significant ornamental trait of plants. The pigment content in colored-leaf plants is affected by specific factors, especially the ratio of carotenoids or anthocyanins to chlorophyll, which leads to red or yellow leaves (Cheng et al. 2015). Developed countries have been engaged in the cultivation and application of colorful plants for more than one 100 years. In the United States, Canada, and other countries, the planting area of colored-leaf trees can reach more than 30% that of all trees (Özbayram & Çiçek 2018).

There are three main categories of pigments: chlorophyll, mainly chlorophyll a and chlorophyll b; carotenoids, mainly carotenoids and lutein; and three flavonoid pigments, also known as anthocyanins (Tanaka et al. 2008). Different pigments appear in different colors; ordinary leaves contain more chlorophyll than carotenoids, so the leaves are always green (Tanaka et al. 2008). Studies have shown that a combination of flavonoids, anthocyanins, and carotenoids can increase leaf color diversity (Luiza Koop et al. 2022). *Ginkgo biloba* L, *Pseudolarix amabilis,* and other leaves contain more carotenoids and appear yellow; anthocyanins, such as in *Acer palmatum* Thunb and *Acer buergerianum* Miq, appear red when their content increases. *Amaranthus tricolor* has leaves from purple to bright red, which is also due to cell senescence and chlorophyll reduction, so that the color of anthocyanins shows (Sharma et al. 2021). Carotenoids and anthocyanins also play an important role in the photoprotection of photosystems, which help plants adapt to the environment and maintain the proper functioning of their internal mechanisms during leaf senescence (Das et al. 2011; Hashimoto et al. 2016).

The massive accumulation of anthocyanins is the basic material for the reddening of colorful plant leaves (Zhang et al. 2019). Many studies have shown that the change in leaf color anthocyanin content in colored-leaf plants is the result of interaction of environmental factors and internal factors (Santos-Buelga et al. 2014). Temperature drop and a shortened photoperiod in autumn and winter often trigger genetic signals for chlorophyll decomposition (Song et al. 2013).

With the decomposition of amino acids, proteins, and other substances in leaves and migration of nutrients in leaves, the chlorophyll content gradually decreases with time, while the anthocyanin content gradually increases, making the leaves red (Li et al. 2018). At the same time, the synthesis and accumulation of anthocyanins is closely related to sugar, protein, chlorophyll, carotenoids, and other substances in plants, and is affected by natural environmental factors such as light time, light intensity, and temperature changes (Zhang et al. 2016).

Anthocyanins are produced by flavonoid compounds through the shikimic acid synthesis pathway, synthesized in the cytoplasm, and eventually stored in plant cell vacuoles (Mishio et al. 2015). This synthetic pathway is an important branch of the plant flavonoid biosynthesis pathway, and it is also one of the most important pathways for the synthesis of plant secondary metabolites (Zhao et al. 2022). The precursors of anthocyanin synthesis are malonic acid monophthalide-CoA and 4-hydroxybenzoic acid-CoA, and the final product anthocyanin is produced by a series of key enzymes, such as chalcone synthase (CHS), chaleone isomerase (CHI), dihydroflavonol reductase (DFR), and glycosyltransferase (GT) (Tanaka et al. 2008). Anthocyanin metabolism is affected by key enzyme genes and transcription factors (Tanaka et al. 2008). Key enzyme genes directly encode the key enzymes required for anthocyanin synthesis (Zhou et al. 2017).

Fraxinus angustifolia is an important afforestation tree species in northern China (Sarfraz et al. 2017). *Fraxinus angustifolia* has an upright body, straight trunk, and luxuriant branches and leaves, and it is an excellent street tree and shade tree, as autumn-colored tree species have become favored for good landscape effects in recent years (Manzanera & Martinez-Chacon 2007; Özbayram & Çiçek 2018). However, research on the change of leaf color has been limited. In this study, transcriptome and metabolome sequencing analysis was performed on the leaves of two critical periods of *Fraxinus angustifolia* to explore the genes related to the leaf coloration of *Fraxinus angustifolia* and their regulatory networks, and then to screen out the genes that may be related to the metabolic process of anthocyanins and leaf color expression. The results of this study provide a theoretical basis for elucidating the mechanism of leaf color formation, and they reveal the role and molecular mechanism of anthocyanins in the leaves of *Fraxinus angustifolia*. This has

important guiding significance for directional cultivation of colored-leaf *Fraxinus* varieties and will also provide important technical support for the study of molecular regulation mechanisms of colored-leaf traits in woody plants.

2. Materials and Methods

2.1 Plant Materials

The leaves of 6-year-old *Fraxinus angustifolia* were collected from Baoding, Hebei Province in October. Then, the leaves of *Fraxinus angustifolia* were quickly frozen in liquid nitrogen bottles, and then transferred to an ultra-low temperature refrigerator for cryopreservation for RNA and metabolite extraction.

2.2 RNA Extraction, Library Construction, and Sequencing

Total RNA was extracted from the sample using TRIzol (Invitrogen, Massachusetts, USA) and purified using a RNeasy column (Qiagen, Dusseldorf, Germany). mRNAs were enriched with Oligo(dT) beads After total RNA extraction. Enriched mRNA fragments were split into short fragments and reverse transcribed into cDNA. The cDNA fragment was then purified using the QiaQuick PCR extraction kit (Qiagen, Venlo, the Netherlands), the ends were repaired, poly(A) was added, and it was connected to the Illumina sequencing adapter. After PCR amplification, sequencing was performed Illumina HiSeqTM 4000.

2.3 Transcriptome Low Quality Data Filtering

FastQC (Brown et al. 2017) software was used to filter the data. The parameters were as follows: a) delete the read containing the adapter; b) remove reads containing more than 10% unknown nucleotides (N); c) remove low quality reads containing more than 50% low quality (q value≤20) bases.

2.4 Transcriptome Data Assembly and Integrity Assessment

The obtained high-quality reads were assembled and annotated using Trinity software (Sewe et al. 2022). Trinity is a modular approach and software package that consists of three parts: Inchworm, Chrysalis, and Butterfly. First, the Inchworm was read by the k-mer method to obtain a linear

130 contigs set. Next, Butterfly aggregated the related contigs corresponding to the unique part of the
131 alternatively spliced transcript or the accessory gene and then constructed a de Bruijn diagram for
132 each related contigs cluster. Finally, Butterfly analyzed the path taken to read and read the pairing
133 in the corresponding de Bruijn diagram, and it then output a linear sequence for each alternatively
134 spliced isoform and script obtained from the fallacy gene. At the same time, BUSCO software was
135 used to evaluate the integrity of the assembly results (Manni et al. 2021).

136 **2.5 Transcriptome Annotation and Differential Gene Analysis**

137 Firstly, the assembled Unigene sequences were compared to the Nr database, KEGG, COG/KOG,
138 and SwissProt by blastx. Based on Nr annotation information, we used Blast2GO software to
139 obtain GO function annotation (Conesa & Gotz 2008). At the same time, the Pfam_Scan program
140 was used to compare with the Pfam database to obtain annotation information related to protein
141 structure. The HMMER program was compared with the SMART database to obtain the
142 annotation information of the protein domain. The PlantTFDB database was compared by
143 hmmscan to obtain transcription factor annotation information (Jin et al. 2017). DESeq2 software
144 was used to analyze differentially expressed genes (DEGs) between groups (Love et al. 2014). The
145 analysis was divided into three parts: a) standardization of readcount; b) calculation of the
146 probability of hypothesis testing (P-value) based on the model; and c) multiple hypothesis test
147 correction to obtain the FDR value (false discovery rate). Based on the results of differential
148 analysis, the genes of FDR were screened as significantly differential genes.

149 **2.6 Gene Function Enrichment Analysis**

150 Gene ontology is an internationally standardized gene function classification system that fully
151 describes the characteristics of genes and their products in organisms. The differential genes are
152 mapped to GO database. AgriGO v2.0 custom tools KOBAS-i was employed to GO and KEGG
153 enrichment analysis, respectively (Tian et al. 2017, Bu et al. 2021).

154 **2.7 Metabolite Extraction**

155 The biomaterial samples, preserved at ultra-low temperatures, were taken out and vacuum freeze-
156 dried. Samples were ground by a grinding machine at 30 Hz for 1.5 min, and 100 mg of powder

was weighed. The powder was extracted overnight at 4℃ with 70% methanol 1.0 ml containing 0.1 mg/l lidocaine as an internal standard, and vortexed three times to make the extraction more sufficient. After extraction, it was centrifuged at 10,000 g for 10 min, the supernatant was aspirated, and the sample was filtered through a microporous membrane (0.22 μm pore size) and stored in a sample bottle for subsequent LC-MS analysis. A quality control sample (QC) was prepared by mixing sample extracts to analyze the repeatability of samples under the same treatment method. Usually, a QC sample was inserted every 10 samples to investigate the repeatability of the analysis process.

2.8 Detection of Metabolites

Conditions for UPLC analysis were as follows: An analytical column Mobile phase: ultrapure water (0.04% acetic acid added), acetonitrile (0.04% acetic acid added); elution gradient: 0 min water/acetonitrile (95:5V/V), 11.0 min 5:95V/V, 12.0 min 5:95V/V, 12.1 min 95:5V/V, and 15.0 min 95:5V/V. Waters ACQUITY UPLC HSS T3 C18 1.8 There was a 0.4 ml/min flow. 40 °C for the column; a 2 l input volume. ESI-QTRAP-MS was used to evaluate the separated samples.

The primary settings for the linear ion trap and triple quadrupole in the API 6500 QTRAP LC/MS/MS system were as follows: collision-activated dissociation (CD) parameter setting of high, electrospray ionization (ESI) temperature of 550 °C, mass spectrometry voltage of 5500 V, curtain gas (CUR) of 25 psi. Each ion pair is scanned in a triple quadrupole (QQQ) using an optimum declustering potential (DP) and collision energy (CE). The software Analyst 1.6.1 was used to process the data.

2.9 Metabolomics Data Analysis

Peaks were manually examined for signal/noise (s/n) > 10, and internal Perl software was used to eliminate duplicate signals brought on by various isotopes, in-source fragmentation, K+, Na+, and NH4+ adduct, and dimerization. This resulted in a matrix with fewer skewed and redundant data. Accurate m/z for each Q1 was collected to make it easier to identify and annotate metabolites. To provide a summary of the metabolite profiles of all samples, total ion chromatograms (TICs) and extracted ion chromatograms (EICs or XICs) of QC samples were exported. Each chromatographic

peak's area was determined. Peak alignment was done between the several samples using the spectral pattern and retention time. Metabolites were identified by searching internal database and public databases (MassBank, KNApSAcK, HMDB, MoToDB, and METLIN) and comparing the m/z values, the RT, and the fragmentation patterns with the standards.

Statistical analysis (PCA analysis and OPLS-DA analysis) used the VIP value and P value to screen differential metabolites. Through inter-group difference analysis, differentially expressed genes were obtained from transcriptome data, differentially expressed metabolites were obtained from metabolomics data, and KEGG enrichment analysis of each group was performed. The correlation analysis of common KEGG pathways was carried out for the differential genes and metabolites between groups.

2. 10 Transcriptome Metabolome O2PLS Model Analysis

Based on all transcriptome and all metabolome data, we performed O2PLS analysis using the OmicsPLS package (Bouhaddani et al. 2018). The O2PLS model is suitable for association analysis of two omics data and can be used for bidirectional modeling and prediction in two data matrices. Through calculation, the O2PLS model decomposes the data of each omics into three parts, namely, the joint part (A part, that is strongly related to another group), the orthogonal part (parts that affect only one group of data and have no effect on another omics data), and the noise part (parts that have no effect on both sets of data). The degree to which each component explains the total variation is expressed as R2, and Higher values represent better model explanatory power. Overfitting and underfitting of models reduce R2.

2.11 Correlation Coefficient Analysis

The Pearson correlation coefficient can be used to measure the relationship between two variables, representing the strength of co-variation between two variables, ranging from -1 to +1. The Pearson coefficient of the expression of all differential genes (the union of differential genes between each comparison group) and the abundance of all differential metabolites (the union of differential metabolites between each comparison group) was calculated to evaluate the correlation between genes and metabolites.

211

212 **3. Results**

213 **3.1 Transcriptome Data filtering and Assembly Analysis**

214 To explore the mechanism of leaf color change in *Fraxinus angustifolia*, we sequenced the *de novo*
215 transcriptomes of *Fraxinus angustifolia* leaves at two different growth stages. Based on the
216 transcriptome data, after quality control, the clean reads of six samples were 5,766,222,300–
217 10,874,368,200, the Q30 base percentage was 93.73–95.20%, and the GC content was 44.98–
218 45.44% (Table S1). The overall sequencing filtering quality was good and could be used for
219 subsequent transcriptome analysis. Through transcriptome assembly, we obtained 74,430
220 Unigenes with a length ranging from 201 bp to 16,676 bp, with an average length of 863 bp (Fig.
221 1B). According to the distribution of Unigene with different lengths after assembly, the length of
222 0–500 bp accounted for about 60% of the total. In the splicing results, the N50 length was 1458
223 bp. The integrity of the assembled transcripts was evaluated by BUSCO (Fig. S1), and the integrity
224 was 75.56%, indicating that the integrity of the transcriptome data was high.

225 **3.2 Functional Annotation Analysis of Unigenes**

226 To explore the gene function of the unigenes, we used four databases (Nr, KEGG, KOG,
227 SwissProt) to annotate the obtained Unigenes (Fig.1C). Finally, 44,746 genes with annotation
228 information were obtained, and 29,684 Unigenes were not annotated, with an annotation ratio of
229 60.12%. Among them, 43,150 Unigenes were annotated in the Nr database, 39,999 Unigenes were
230 annotated in the KEGG database, 24,903 Unigenes were annotated in the KOG database, and
231 30,315 Unigenes were annotated in the Swissprot database. Among them, 20,659 genes had
232 annotation information in four databases.

233 **3.3 Gene Differential Expression Analysis of Transcriptome Data**

234 In order to ensure the accuracy of subsequent analysis, we first corrected the sequencing depth,
235 and then corrected the length of the gene or transcript to obtain the RPKM value of the gene, and
236 then performed the subsequent analysis. Based on the RPKM value of each gene, the expression
237 distribution was calculated. As can be seen from the distribution, gene expression is more uniform

between different samples, meaning that these results can be used for subsequent differential gene analysis (Fig. 2A). Principal component analysis (PCA) and pearson correlation coefficient analysis clustered the samples based on the expression level. Through PCA analysis between these samples, results showed that three biological replicates of samples from different periods were distributed in the same quadrant (Fig. 2B). Pearson correlation coefficient analysis showed the correlation of each two samples was above 0.85 (Fig. S2), indicating that the sample repeatability was good. A total of 5827 DEGs were found between stage 1 (green leaf stage) and stage 2 (red purple leaf stage), of which 2249 were upregulated and 3578 were downregulated (Figure 2C and 2D).

3.4 Functional Analysis of Differential Genes

To elucidate the function of differentially expressed genes in leaf color change of *Fraxinus angustifolia*, GO enrichment analysis of differentially expressed genes was performed. Based on the results of GO functional enrichment analysis, the differentially expressed genes in stage 1 (green leaf stage) and stage 2 (red purple leaf stage) were mainly enriched in flavonoid anabolism, programmed cell death, phenylpropanoid anabolism, chloroplast quinone metabolism, pigment metabolism, carotene metabolism, polysaccharide metabolism, terpenoid anabolism, jasmonic acid metabolism, fatty acid anabolism, photosynthesis, secondary metabolite biosynthesis, pigment accumulation, and other biological processes (Fig. 3A). In the molecular function classification, the differentially expressed genes were significantly enriched in terms of oxidoreductase activity, monooxygenase activity, hydrolase activity, transferase activity, tetrapyrrole binding, nucleic acid binding transcription factor activity, carotenoid dioxygenase activity, antioxidant activity, fatty acid synthase activity, galactosyltransferase activity, and iron ion binding (Fig. 3B). In the classification of cell composition, GO terms were mainly enriched in the thylakoid, chloroplasts, photosynthetic membranes, cell wall, peroxisome, photosystem, light-harvesting complex, thylakoid membrane, etc (Fig. 3C).

Then, KEGG pathway enrichment analysis was performed with differentially expressed genes to analyze the key enriched metabolic pathways in leaf color change of *Fraxinus*

angustifolia. Based on the results of KEGG metabolic pathway enrichment analysis, we found that the differentially expressed genes between stage 1 (green leaf) and stage 2 (red-purple leaf) were significantly enriched in metabolic pathways such as secondary metabolite biosynthesis, phenylpropanoid biosynthesis, riboflavin metabolism, flavonoid biosynthesis, anthocyanin biosynthesis, carotene biosynthesis, terpenoid biosynthesis, steroid biosynthesis, and thiamine metabolism (Fig. 4). The functional enrichment analysis indicated that differentially expressed genes played an important role in leaf color change of *Fraxinus angustifolia*.

3.5 Analysis of Differential Metabolites in Leaves of *Fraxinus angustifolia* at Different Stages

Leaves show different colors mainly due to internal pigment-related components and content. The leaves of different periods of *Fraxinus angustifolia* have significant color differences in phenotype (Fig. 1A). Through the determination of the metabolome, the leaves of *Fraxinus angustifolia* in these two periods can be divided into two groups with significant differences according to PCA analysis, based on the content of their internal metabolites (Fig. 5A). To further identify the metabolites associated with leaf color difference, we used orthogonal partial least squares discriminant analysis (OPLS-DA) to explore the differential metabolites between different colors.

Cross-validation and ranking validation showed that the OPLS-DA model of each comparison group was meaningful (Fig. 5B). Then, we obtained the relevant load diagram by OPLS-DA analysis (Fig. 5B and S3). The farther the metabolites from the origin in the OPLS-DA loading plot, the greater their contribution to the separation of leaf color categories of *Fraxinus angustifolia*. Organic acids, lipids, and flavonols are the main substances that cause differences in the metabolome level of *Fraxinus angustifolia* at different stages, while flavonols are the main factors that cause differences in leaf color at different stages.

Furthermore, according to the criteria for identifying differential metabolites (VIP>1, t-test P-value<0.05), we identified a total of 118 compounds. Compared with stage 1 (green leaf stage) and stage 2 (red purple leaf stage), with 40 kinds of material accumulation, there were 13 kinds of flavonoid compounds, including 8 kinds of flavonols and 4 kinds of flavonoids (Fig. 5C). Therefore, we speculate that flavonoid compounds may be the main factor causing the change of

leaf color of *Fraxinus angustifolia*.

Based on the metabolic pathway enrichment analysis of differential metabolites, we found that the differential metabolites in stage 1 (green leaf stage) and stage 2 (red purple leaf stage) were mainly enriched in secondary metabolite biosynthesis, aminoacyl-tRNA biosynthesis, 2-oxocarboxylic acid metabolism, glucosinolate biosynthesis, carbapenem biosynthesis, cyanoamino acid metabolism, phenylpropanoid biosynthesis, ubiquinone and other terpenoid quinone biosynthesis, riboflavin metabolism, porphyrin and chlorophyll metabolism, fatty acid biosynthesis, the TCA cycle, tryptophan metabolism, niacin and nicotinamide metabolism, and other metabolic pathways (Fig. 5D). Based on the above results, we found that the pathways related to anthocyanin synthesis were significantly enriched, which provided an important basis for subsequent analysis.

3.6 Association Analysis of DEGs and Metabolites Between the Green and Red-Purple Leaf Stage

To calculate the gene expression and metabolite abundance, we used pearson correlation coefficient to evaluate the correlation between genes and metabolites in stage 1 and stage 2. Then, we performed a loading plot analysis on the association between genes and metabolites to analyze the contribution of each variable (metabolites, genes) to the differences between groups based on the O2PLS model analysis. The loading value represents the explanatory power of the variables (metabolites, genes) in each component (the contribution to the difference between groups), and the positive or negative value of the loading value indicates a positive or negative correlation with another group. The larger the absolute value of the load value, the stronger the correlation. By comparing and analyzing the correlation between differential genes and metabolites in the two periods, we found that they were located at both ends of the ordinate; that is, the absolute values were very large (Figure 2), indicating that the data in this part were highly correlated, which provided a reliable basis for further analysis. Through the analysis of the most relevant top 25 results in the association analysis results, we found that the metabolites in stage 1 (green leaf stage) and stage 2 (red-purple leaf stage) were amino acids and their derivatives, sesquiterpenes,

dihydroflavonols, flavonoids, flavonols, triterpenoids, vitamins, coumarins, phenolic acids, lignans, and organic acids, respectively. The genes associated with them were MYB transcription factors, DNA glycosylase superfamily proteins, NAC domain-containing proteins, etc. (Fig. 6A).

Through the network association analysis of genes or metabolites in important association positions, we screened out the key associated genes and metabolites, and focused on the analysis of the results of the top 250 (Figure 6B). The core metabolites were flavonols (Lmmp002334, Lmyp004444, Hmpp003242), alkaloids (pmp001275), free fatty acids (pmp001276), phenolic acids (mws0885, mws0183), flavonoids (Hmpp003270), amino acids and their derivatives (pme0193), and organic acids (mws0639). The genes related to them were various transcription factors (*MYB, NAC, bHLH, WRKY, TCP*, etc.) and various enzymes.

3.7 Screening of Anthocyanin Synthesis-Related Genes

Based on the above transcriptome and metabolomics data analysis results, we found that differential genes or substances were mainly concentrated in the anthocyanin synthesis pathway. Therefore, we analyzed the genes related to the anthocyanin synthesis pathway. Through the analysis of expression differences, we screened seven genes with higher expression levels in the red and purple leaf stage of the leaves of *Fraxinus angustifolia* (Figure 7A and 7B), which were *Unigene0029203*(*4CL3*), *Unigene0037675*(*PAL*), *Unigene0061770*(*DFRA*), *Unigene0009106*(*CHI*), *Unigene0010118*(*PAL*), *Unigene0037674*(*PALA*), and *Unigene0002804*(*CHI3*). The significant differential expression of these genes in two different periods may be closely related to the change of leaf color.

Anthocyanins are not stable in plant cells, and they are generally modified (methylation, acylation, glycosylation, etc.) to form more stable and richer anthocyanidin, which directly determine the color of plant tissues. We further analyzed the genes related to the anthocyanidin biosynthesis pathway in the transcriptome data and found that two Unigenes in the anthocyanidin biosynthesis pathway were upregulated at the red-purple leaf stage (Fig. 7C and 7D), one of which was *Bronze-1* (*BZ1*, *anthocyanidin 3-O-glucosyltransferase gene*) and the other gene was *UGT79B1* (*anthocyanidin 3-O-glucoside 2''-O-xylosyltransferase gene*) (Fig. S4). These two genes

346 showed significant expression differences in different samples, indicating that they may play an
347 important role in leaf color change.
348
349 **4. Discussion**
350 Environmental changes can lead to changes in the color of plant leaves. Usually, changes in
351 external conditions such as temperature, light, humidity, stress environment, reproductive growth,
352 and other external conditions can induce plant leaves to produce corresponding signals, thereby
353 regulating the expression of leaf color change-related genes and the occurrence of related
354 physiological processes, so that leaves show different colors (Sork et al. 2010). Studying the
355 physiological and molecular regulation mechanism of leaf discoloration of color-leafed plants,
356 exploring the related genes involved in leaf color regulation, and clarifying their functions and
357 mechanisms will lay a solid theoretical foundation for enriching the landscape and provide
358 important theoretical guidance for the application and promotion of color-leafed tree species (Zhao
359 et al. 2020).
360 Pigments in plants mainly include chlorophyll, carotenoids, flavonoids, and anthocyanins
361 (Zhao et al. 2020). Red is mainly derived from the composition and content of flavonoids.
362 Flavonoids are mainly anthocyanins and belong to the more important secondary metabolites,
363 mainly cyanidin aglycone, delphinidin aglycone, pelargonidin aglycone, peoniflorin aglycone,
364 petunidin, and malvidin 6 anthocyanins derived from aglycone. When the anthocyanin content is
365 high, the leaves appear red (Kong et al. 2003). New red leaves of tree peony are due to increased
366 anthocyanin content and decreased chlorophyll and carotenoid content (Luo et al. 2017). The
367 anthocyanin content in young leaves of jatropha is 11 times the chlorophyll content, mostly in
368 epidermal cells, making leaves red (Ranjan et al. 2014). In this study, we performed transcriptome
369 analysis of the leaves of *Fraxinus angustifolia* and found that 5827 genes were differentially
370 expressed at two different developmental stages. GO enrichment analysis showed that the
371 differentially expressed genes in the three periods were significantly enriched in flavonoid
372 anabolism, phenylpropanoid anabolism, chlorophyll quinone metabolism, pigment metabolism,

carotene metabolism, polysaccharide metabolism, terpenoid anabolism, jasmonic acid metabolism, fatty acid anabolism, photosynthesis, secondary metabolite biosynthesis, pigment accumulation, and other terms, indicating that the differentially expressed genes enriched in these terms may play an important role in leaf color change of *Fraxinus angustifolia*.

At the same time, KEGG metabolic pathway enrichment analysis showed that the differential genes were significantly enriched in the phenylpropanoid biosynthesis pathway, which is responsible for the anabolism of flavonoids, such as anthocyanins, flavonols, and flavanols (Liu et al. 2021). In many plants, the flavonoid pathway exists in colored tissues and is the main cause of color changes, such as *Tulipa gesneriana* (Guo et al. 2022), *Prunus persica*(Ravaglia et al. 2013), etc. Through gene function annotation, it is valuable to understand the accumulation of anthocyanins, and it is also important for further study of functional genes (Santos-Buelga et al. 2014; Belwal et al. 2020). In this study, KEGG enrichment analysis identified two important pathways related to anthocyanin biosynthetic metabolism; the flavonoid biosynthesis pathway, flavonoid and flavonol biosynthesis pathways, and anthocyanin metabolic pathways related to anthocyanin substrate synthesis. Therefore, these genes may have a direct correlation with the synthesis of anthocyanin in *Fraxinus angustifolia*.

Plant metabolomics enables qualitative and quantitative analysis of small molecule metabolites and understanding of their synthesis and metabolic patterns (Schauer & Fernie 2006; Wang & Kadarmideen 2020). At present, the results show that the study of plant metabolites is important for the identification of metabolites, variety differentiation and molecular breeding (Rizzato et al. 2017; Zhou et al. 2017). The formation of plant color is closely related to the anabolism of flavonoids (Zhang et al. 2017). In the metabolomic analysis of flavonoids in white and purple flowers of Phalaenopsis, 142 different flavonoid-related metabolites were identified, proving that the major anthocyanin is a cyanide derivative (Meng et al. 2020). In this study, we determined the metabolites of the leaves of *Fraxinus angustifolia*, and OPLS-DA analysis showed that flavonols were the main factors for the difference of leaf color in different periods of *Fraxinus angustifolia*, and 118 compounds were identified. It is speculated that flavonoid compounds may

be the main factor causing leaf color changes. The metabolic pathway enrichment analysis of differential metabolites showed that the metabolic pathways related to anthocyanins synthesis were significantly enriched, which provided an important basis for downstream analysis.

It has been confirmed in plants that transcription factor families, such as MYB and bHLH, are essential for inducing anthocyanin biosynthesis pathways (Allan et al. 2008; Tanaka et al. 2008). For example, mutations in these regulatory genes usually lead to color changes in apple, *Actinidia chinensis* purple cauliflower (*Brassica oleracea* L.), *Dahlia pinnata,* and tomatoes (Butelli et al. 2008; Chiu et al. 2010; Ohno et al. 2011; Wang et al. 2019; Sun et al. 2021). We further performed correlation analysis of transcriptome and metabolome data, and we found that there were multiple common metabolic pathways in stage 1 (green leaf stage) and stage 2 (red-purple leaf stage). Based on the O2PLS model analysis, we found that the flavonoid compounds in the substances closely related to the genes in two different periods had a high proportion and were at the core association position of the metabolite network diagram. We found that flavonoid compounds are closely related to transcription factors, such as phenolic acids (mws0183, mws0885, mws1195, mws1200, and pmb2795), flavonoids (HJAP064, HJAP065, Hmpp003270, Xmyp005261, and Zmhp005139), and flavonols (Hmpp003242, Lmdp003808, Lmmp002334, Lmyp004444, and pmp001311). Association analysis results showed that more than 200 genes were closely related to these metabolites, among which transcription factors such as MYB, bHLH, and NAC were closely related to these metabolites. Therefore, these results indicated that transcription factors play an important role in the change of flavonoid compounds during the change of leaf color in autumn. In addition, we also screened nine differentially expressed genes related to anthocyanins. Transcriptome data and qRT-PCR results showed that these genes had significant expression differences in samples of different periods of *Fraxinus angustifolia*. We predicted that they were the key structural genes of anthocyanin biosynthesis and the main factors regulating leaf color changes, which provided an important basis for further analysis of the molecular regulation mechanism of leaf coloration of *Fraxinus angustifolia*.

427

Author contributions

Conceived and designed the experiments: Y. Li, K. Zhang, Z. Wang. Performed the experiments: Y. Wang, X. Che, J. Zhen, G. Zhang, H. Yang, J. Wen. Data analysis: Y. Wang, X. Che, J. Zhen, J. Wang, J. Wang, B. He, A. Yu. Wrote the paper: Y. Li, K. Zhang, Z. Wang.

Funding

This work was supported by Key project of Education Department of Hebei Province (ZD2022009), National Key Research and Development Project (2020YFD1000700), and Funding Program for Visiting Scholar Abroad (C20200334).

Data Availability

The RNA-seq datasets generated in this study are available in National Genomics Data Center (NGDC) databases, under Bioproject accession number PRJCA014072 that is publicly accessible at https://ngdc.cncb.ac.cn/.

Competing Interests

The authors declare no conflict of interest.

446 **Figure legends**

447 **Figure 1.** Statistical analysis of transcriptome data. (A) Leaf states at stage 1 and stage 2. (B)
448 Length distribution of Unigene in transcriptome assembly. (C) Venn diagram of Unigene
449 annotation in the Nr, KEGG, KOG, and SwissProt databases.

450

451 **Figure 2.** Expression analysis of Unigenes in stage 1 and stage 2. (A) Expression level analysis of
452 different biological replicates in stage 1 and stage 2. (B) Principal component analysis (PCA) of
453 expression datasets in stage 1 and stage 2. (C) Volcano-plots of differentially expressed genes in
454 stage 1 and stage 2. (D) Heatmap of differentially expressed genes in stage 1 and stage 2. The
455 numbers ranging from -2 to 2 represented the expression level from low to high based Z-score
456 method.

457

458 **Figure 3.** GO function enrichment analysis of differentially expressed genes. (A) Biological
459 progress. (B) Molecular function. (C) Cellular component. Fisher's exact test, adjusted P-value
460 was performed by the Benjamini-Yekutieli method (FDR < 0.05).

461

462 **Figure 4.** KEGG enrichment analysis of differentially expressed genes of stage 1 vs stage 2. The
463 abscissa indicates the degree of significant enrichment.

464

465 **Figure 5.** Analysis of leaf metabolites in different periods. (A) Differential metabolite analysis on
466 the basis of principal component analysis (PCA) between stage 1 and stage 2. (B) OPLS-DA model
467 plots for the comparison group stage 1 vs stage 2. (C) Differential metabolite heatmap of stage 1
468 vs stage 2, metabolites ID can be found in Table S2. (D) Analysis of common KEGG metabolic
469 pathways between stage 1 and stage 2.

470

471 **Figure 6.** Association analysis of metabolomics and transcriptomics data. (A) Divergent gene and
472 metabolite loading map of stage 1 and stage 2. Each point in the figure represents a gene or a

metabolite. The abscissa is the first-dimensional coordinate of the joint part, and the ordinate is the second-dimensional coordinate of the joint part. The greater the absolute value of the element in the coordinate, the greater the correlation between this element and another group. (B) Network diagram of the correlation between differential gene expression and differential metabolite abundance, top 250 pairs of genes and metabolites (with the absolute pearson correlation > 0.5) were applied for metabolite-transcript network analysis.

Figure 7. Expression level of anthocyanin and synthesis pathway-related genes. (A) Transcriptome data of anthocyanin synthesis pathway-related genes. (B) qRT-PCR results of anthocyanin synthesis pathway-related genes in different stages. Error bar represents the mean ± SD of three independent biological replicates. (C) Transcriptome data of anthocyanidin synthesis pathway-related genes *BZ1* and *UGT79B1*. (D) qRT-PCR results of *BZ1* and *UGT79B1* in two stages. Error bar represents the mean ± SD of three independent biological replicates.

Figure S1. Integrity evaluation of *Fraxinus angustifolia* transcriptome data in stage 1 and stage 2.

Figure S2. Pearson correlation coefficient analysis based on the expression level of stage 1 and stage 2.

Figure S3. OPLSDA loadings plot of *Fraxinus angustifolia* metabolites in stage 1 and stage 2.

Figure S4. Anthocyanin biosynthesis diagrams based on differentially expressed genes in stage 1 and stage 2.

Table S1. Statistics of transcriptome clean data.

Table S2. Name, type and expression of metabolites in stage 1 and stage 2, values of stage 1 and stage 2 represent expression data of metabolites.

References

Allan AC, Hellens RP, and Laing WA. 2008. MYB transcription factors that colour our fruit. *Trends in Plant Science* **13:**99-102 DOI 10.1016/j.tplants.2007.11.012

Belwal T, Singh G, Jeandet P, Pandey A, Giri L, Ramola S, Bhatt ID, Venskutonis PR, Georgiev MI, Clement C, and Luo Z. 2020. Anthocyanins, multi-functional natural products of industrial relevance: Recent biotechnological advances. *Biotechnology Advances* **43:**107600 DOI 10.1016/j.biotechadv.2020.107600

Bouhaddani SE, Uh HW, Jongbloed G, Hayward C, Klaric L, Kielbasa SM, and Houwing-Duistermaat J. 2018. Integrating omics datasets with the OmicsPLS package. *BMC Bioinformatics* **19:**371 DOI 10.1186/s12859-018-2371-3

Brown J, Pirrung M, and McCue LA. 2017. FQC Dashboard: integrates FastQC results into a web-based, interactive, and extensible FASTQ quality control tool. *Bioinformatics* **33:**3137-3139 DOI 10.1093/bioinformatics/btx373

Butelli E, Titta L, Giorgio M, Mock HP, Matros A, Peterek S, Schijlen EG, Hall RD, Bovy AG, Luo J, and Martin C. 2008. Enrichment of tomato fruit with health-promoting anthocyanins by expression of select transcription factors. *Nature Biotechnology* **26:**1301-1308 DOI 10.1038/nbt.1506

Cheng J, Liao L, Zhou H, Gu C, Wang L, and Han Y. 2015. A small indel mutation in an anthocyanin transporter causes variegated colouration of peach flowers. *Journal of Experimental Botany* **66** DOI 10.1093/jxb/erv419

Chiu LW, Zhou X, Burke S, Wu X, Prior RL, and Li L. 2010. The purple cauliflower arises from activation of a MYB transcription factor. *Plant Physiology* **154:**1470-1480 10.1104/pp.110.164160

Conesa A, and Gotz S. 2008. Blast2GO: A comprehensive suite for functional analysis in plant genomics. *International Journal of Plant Genomics* **2008:**619832 DOI 10.1155/2008/619832

Das PK, Geul B, Choi SB, Yoo SD, and Park YI. 2011. Photosynthesis-dependent anthocyanin pigmentation in Arabidopsis. *Plant signaling & behavior* **6:**23-25 DOI 10.4161/psb.6.1.14082

Guo X, Fu X, Li X, and Tang D. 2022. Effect of flavonoid dynamic changes on flower coloration of *Tulipa gesneiana* 'Queen of Night' during clower development. *Horticulturae* **8**, 510 DOI 10.3390/horticulturae8060510

Hashimoto H, Uragami C, and Cogdell RJ. 2016. Carotenoids and Photosynthesis. *Subcellular Biochemistry* **79:**111-139 DOI 10.1007/978-3-319-39126-7_4

Jin J, Tian F, Yang DC, Meng YQ, Kong L, Luo J, and Gao G. 2017. PlantTFDB 4.0: toward a central hub for transcription factors and regulatory interactions in plants. *Nucleic Acids Research* **45:** D1040-D1045 DOI 10.1093/nar/gkw982

Kong JM, Chia LS, Goh NK, Chia TF, and Brouillard R. 2003. Analysis and biological activities of anthocyanins. *Phytochemistry* **64:**923-933 DOI 10.1016/s0031-9422(03)00438-2

Li Y, He N, Hou J, Xu L, Liu C, Zhang J, Wang Q, Zhang X, and Wu X. 2018. Factors influencing leaf chlorophyll content in natural forests at the biome scale. *Frontiers in Ecology and Evolution* **6:**64 DOI 10.3389/fevo.2018.00064

Liu W, Feng Y, Yu S, Fan Z, Li X, Li J, and Yin H. 2021. The flavonoid biosynthesis network in plants. *International Journal Of Molecular Sciences* **22:**12824 DOI 10.3390/ijms222312824

Love MI, Huber W, and Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* **15:**550 DOI 10.1186/s13059-014-0550-8

Luiza Koop B, Nascimento da Silva M, Diniz da Silva F, Thayres Dos Santos Lima K, Santos Soares L, Jose de Andrade C, Ayala Valencia G, and Rodrigues Monteiro A. 2022. Flavonoids, anthocyanins, betalains, curcumin, and carotenoids: Sources, classification and enhanced stabilization by encapsulation and adsorption. *Food Research International* **153:**110929 DOI 10.1016/j.foodres.2021.110929

Luo J, Shi Q, Niu L, and Zhang Y. 2017. Transcriptomic analysis of leaf in tree peony reveals differentially expressed pigments genes. *Molecules* **22** DOI 10.3390/molecules22020324

Manni M, Berkeley MR, Seppey M, and Zdobnov EM. 2021. BUSCO: Assessing Genomic Data Quality and Beyond. *Current Protocol* **1:**e323 DOI 10.1002/cpz1.323

Manzanera JA, and Martinez-Chacon MF. 2007. Ecophysiological competence of Populus alba

552 L., *Fraxinus angustifolia* Vahl., and *Crataegus monogyna* Jacq. used in plantations for the
553 recovery of riparian vegetation. *Environment Management* **40:**902-912 DOI 10.1007/s00267-007-
554 9016-z
555 **Meng X, Li G, Gu L, Sun Y, Li Z, Liu J, Wu X, Dong T, and Zhu M. 2020.** Comparative
556 metabolomic and transcriptome analysis reveal distinct flavonoid biosynthesis regulation between
557 petals of white and purple *Phalaenopsis amabilis*. *Journal of Plant Growth Regulation* **39:**823-
558 840 DOI 10.1007/s00344-019-10025-y
559 **Mishio T, Takeda K, and Iwashina T. 2015.** Anthocyanins and other flavonoids as flower
560 pigments from eleven *Centaurea* species. *Nature Product Communacation* **10:**447-450 DOI
561 10.1177/1934578X15010003
562 **Ohno S, Hosokawa M, Hoshino A, Kitamura Y, Morita Y, Park K-I, Nakashima A, Deguchi**
563 **A, Tatsuzawa F, Doi M, Iida S, and Yazawa S. 2011.** A bHLH transcription factor, DvIVS, is
564 involved in regulation of anthocyanin synthesis in dahlia (*Dahlia variabilis*). *Journal of*
565 *Experimental Botany* **62:**5105-5116 DOI 10.1093/jxb/err216
566 **Özbayram AK, and Çiçek E. 2018.** Thinning experiments in narrow-leaved ash (*Fraxinus*
567 *angustifolia* Vahl.) plantations: 10-year results. *New Forests* **49:**585-598 DOI 10.1007/s11056-
568 018-9642-8
569 **Ranjan S, Singh R, Singh M, Pathre UV, and Shirke PA. 2014.** Characterizing photoinhibition
570 and photosynthesis in juvenile-red versus mature-green leaves of *Jatropha curcas* L. *Plant*
571 *Physiology and Biochemistry* **79:**48-59 DOI 10.1016/j.plaphy.2014.03.007
572 **Ravaglia D, Espley RV, Henry-Kirk RA, Andreotti C, Ziosi V, Hellens RP, Costa G, and**
573 **Allan AC. 2013.** Transcriptional regulation of flavonoid biosynthesis in nectarine (*Prunus*
574 *persica*) by a set of R2R3 MYB transcription factors. *BMC Plant Biology* **13:**68 DOI
575 10.1186/1471-2229-13-68
576 **Rizzato G, Scalabrin E, Radaelli M, Capodaglio G, and Piccolo O. 2017.** A new exploration
577 of licorice metabolome. *Food Chemistry* **221:**959-968 DOI 10.1016/j.foodchem.2016.11.068
578 **Santos-Buelga C, Mateus N, and De Freitas V. 2014.** Anthocyanins, plant pigments and beyond.

Journal of Agricultural and Food Chemistry **62:**6879-6884 DOI 10.1021/jf501950s

Sarfraz I, Rasul A, Jabeen F, Younis T, Zahoor MK, Arshad M, and Ali M. 2017. Fraxinus: a plant with versatile pharmacological and biological activities. *Evidence-based Complementary and Alternative Medicine* **2017:**4269868 DOI 10.1155/2017/4269868

Schauer N, and Fernie AR. 2006. Plant metabolomics: towards biological function and mechanism. *Trends in Plant Science* **11:**508-516 DOI 10.1016/j.tplants.2006.08.007

Sewe SO, Silva G, Sicat P, Seal SE, and Visendi P. 2022. Trimming and validation of illumina short reads using trimmomatic, trinity assembly, and assessment of RNA-Seq data. *Methods Molecular Biology* **2443:**211-232 DOI 10.1007/978-1-0716-2067-0_11

Sharma A, Mazumdar B, and Keshav A. 2021. Valorization of unsalable *Amaranthus tricolour* leaves by microwave-assisted extraction of betacyanin and betaxanthin. *Biomass Conversion and Biorefinery* DOI 10.1007/s13399-020-01267-y

Song YH, Ito S, and Imaizumi T. 2013. Flowering time regulation: photoperiod- and temperature-sensing in leaves. *Trends in Plant Science* **18:**575-583 10.1016/j.tplants.2013.05.003

Sork VL, Davis FW, Westfall R, Flint A, Ikegami M, Wang H, and Grivet D. 2010. Gene movement and genetic association with regional climate gradients in California valley oak (*Quercus lobata* Nee) in the face of climate change. *Molecular Ecology* **19:**3806-3823 DOI 10.1111/j.1365-294X.2010.04726.x

Sun C, Wang C, Zhang W, Liu S, Wang W, Yu X, Song T, Yu M, Yu W, and Qu S. 2021. The R2R3-type MYB transcription factor MdMYB90-like is responsible for the enhanced skin color of an apple bud sport mutant. *Horticulture Research* **8:**156 DOI 10.1038/s41438-021-00590-3

Tanaka Y, Sasaki N, and Ohmiya A. 2008. Biosynthesis of plant pigments: anthocyanins, betalains and carotenoids. *Plant J* **54:**733-749 DOI 10.1111/j.1365-313X.2008.03447.x

Tian T, Liu Y, Yan H, You Q, Yi X, Du Z, Xu W, and Su Z. 2017. AgriGO v2.0: a GO analysis toolkit for the agricultural community, 2017 update. *Nucleic Acids Research*, **45**, W122-W129 DOI 10.1093/nar/gkx382

Wang L, Tang W, Hu Y, Zhang Y, Sun J, Guo X, Lu H, Yang Y, Fang C, Niu X, Yue J, Fei Z, and Liu Y. 2019. A MYB/bHLH complex regulates tissue-specific anthocyanin biosynthesis in the inner pericarp of red-centered kiwifruit *Actinidia chinensis* cv. Hongyang. *The Plant Journal* **99:**359-378 DOI 10.1111/tpj.14330

Wang X, and Kadarmideen HN. 2020. Metabolite genome-wide association study (mGWAS) and gene-metabolite interaction network analysis reveal potential biomarkers for feed efficiency in pigs. *Metabolites* **10:**201 DOI 10.3390/metabo10050201

Wang ZX, Yu YF, Chen L, Qin HY, Liu YX, Ai J, Zhao Y, Yang YM, Li XY, Wang GF, Sun D, Wang CW, and Fan ST. 2016. Advances in leaf pigment composition, structure and photosynthetic characteristics of colored-leaf plants. *Plant Physiology Journal* **52:**1-7 DOI 10.13592/j.cnki.ppj.2015.0490

Zhang DW, Yuan S, Xu F, Zhu F, Yuan M, Ye HX, Guo HQ, Lv X, Yin Y, and Lin HH. 2016. Light intensity affects chlorophyll synthesis during greening process by metabolite signal from mitochondrial alternative oxidase in Arabidopsis. *Plant Cell & Environment* **39:**12-25 DOI 10.1111/pce.12438

Zhang Q, Liu M, and Ruan J. 2017. Metabolomics analysis reveals the metabolic and functional roles of flavonoids in light-sensitive tea leaves. *BMC Plant Biology* **17:**64 DOI 10.1186/s12870-017-1012-8

Zhang Q, Zhai J, Shao L, Lin W, and Peng C. 2019. Accumulation of anthocyanins: an adaptation strategy of *Mikania micrantha* to low temperature in winter. *Frontiers in Plant Science* **10:**1049 DOI 10.3389/fpls.2019.01049

Zhao M-H, Li X, Zhang X-X, Zhang H, and Zhao X-Y. 2020. Mutation mechanism of leaf color in plants: a review. *Forests* **11:**851 DOI 10.3390/f11080851

Zhao X, Zhang Y, Long T, Wang S, and Yang J. 2022. Regulation mechanism of plant pigments biosynthesis: anthocyanins, carotenoids, and betalains. *Metabolites* **12** DOI 10.3390/metabo12090871

Zhao Y, and Kong L. 2021. KOBAS-i: intelligent prioritization and exploratory visualization of

biological functions for gene enrichment analysis. *Nucleic Acids Research*, **49**, W317-W325 DOI 10.1093/nar/gkab447

Zhou F, Peng J, Zhao Y, Huang W, Jiang Y, Li M, Wu X, and Lu B. 2017. Varietal classification and antioxidant activity prediction of Osmanthus fragrans Lour. flowers using UPLC-PDA/QTOF-MS and multivariable analysis. *Food Chemistry* **217:**490-497 DOI 10.1016/j.foodchem.2016.08.125

Figure 1

Figure 1

Figure 1. Statistical analysis of transcriptome data. (A) Leaf states at stage 1 and stage 2. (B) Length distribution of Unigene in transcriptome assembly. (C) Venn diagram of Unigene annotation in the Nr, KEGG, KOG, and SwissProt databases.

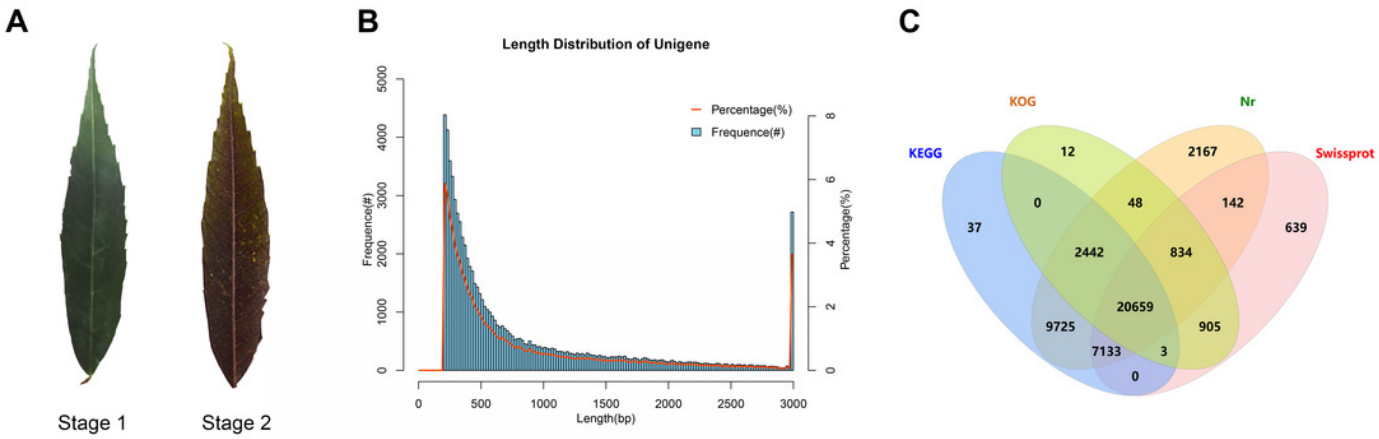


Figure 2

Figure 2

Figure 2. Expression analysis of Unigenes in stage 1 and stage 2. (A) Expression level analysis of different biological replicates in stage 1 and stage 2. (B) Principal component analysis (PCA) of expression datasets in stage 1 and stage 2. (C) Volcano-plots of differentially expressed genes in stage 1 and stage 2. (D) Heatmap of differentially expressed genes in stage 1 and stage 2. The numbers ranging from -2 to 2 represented the expression level from low to high based Z-score method.

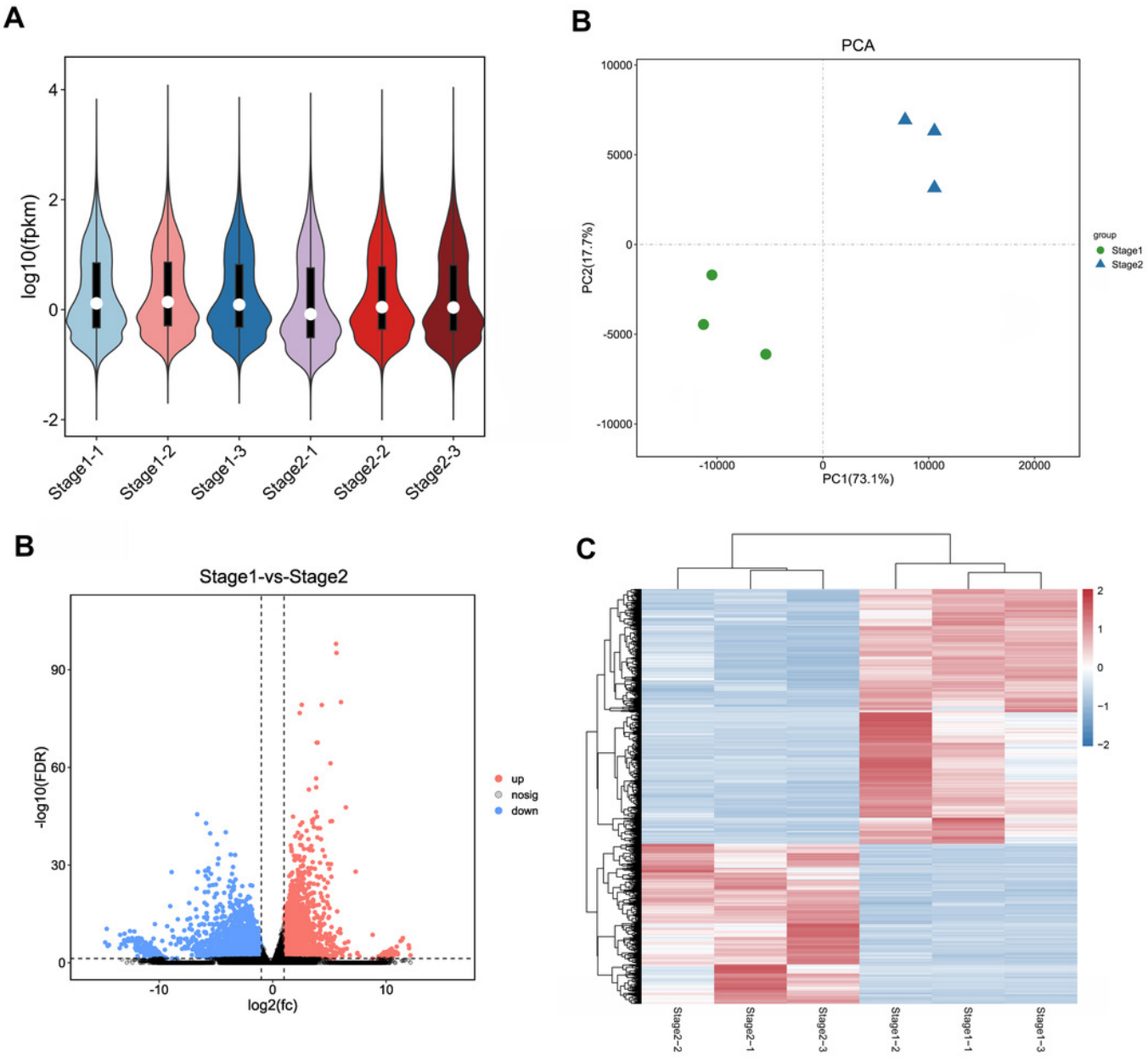

A
log10(fpkm)
Stage1-1
Stage1-2
Stage1-3
Stage2-1
Stage2-2
Stage2-3
B
PCA
PC2(17.7%)
PC1(73.1%)
group
Stage1
Stage2
B
Stage1-vs-Stage2
-log10(FDR)
log2(fc)
up
nosig
down
C
Stage2-2
Stage2-1
Stage2-3
Stage1-2
Stage1-1
Stage1-3

Figure 3

Figure 3

Figure 3. GO function enrichment analysis of differentially expressed genes. (A) Biological progress. (B) Molecular function. (C) Cellular component. Fisher's exact test, adjusted P-value was performed by the Benjamini-Yekutieli method (FDR < 0.05).

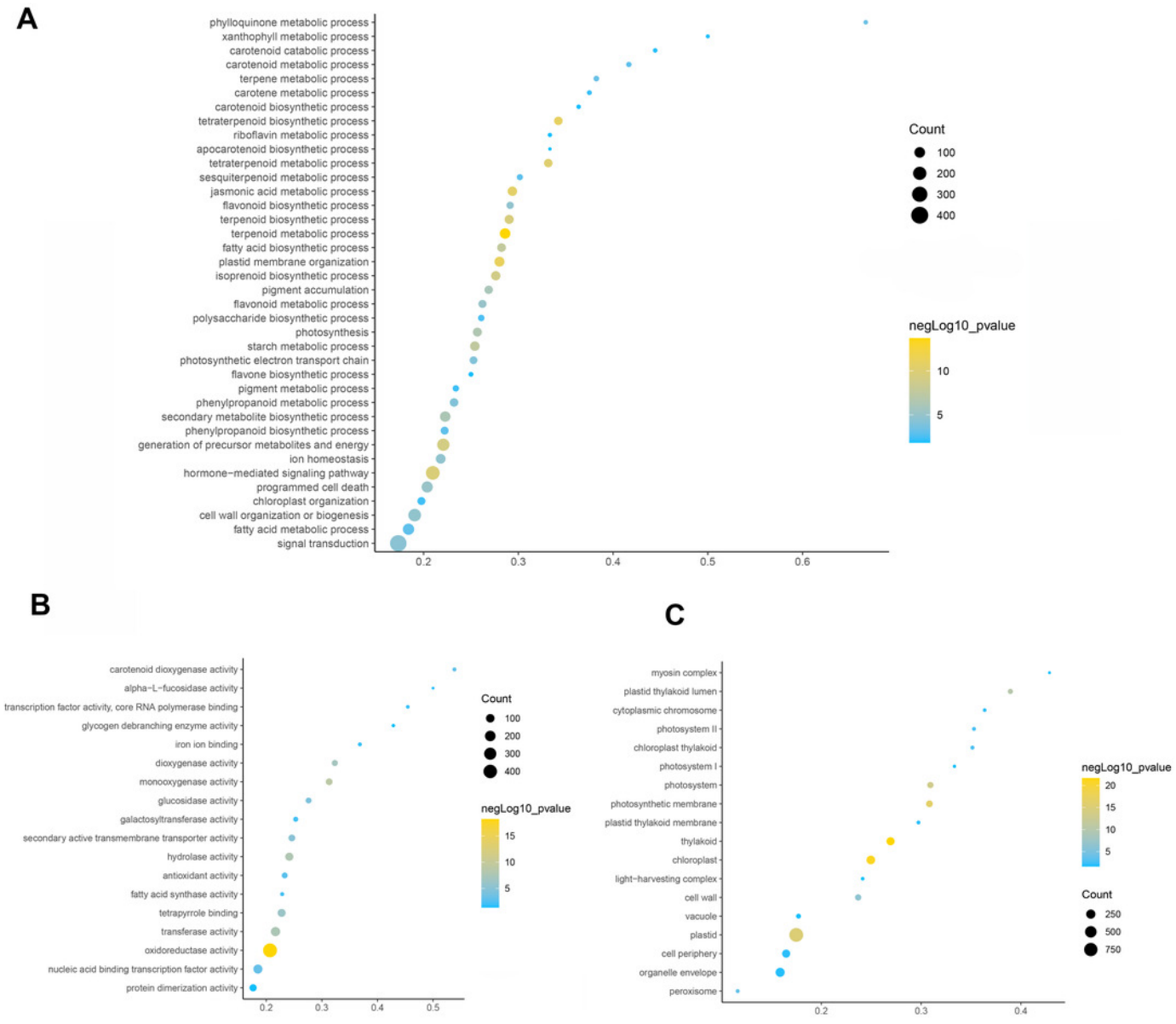


Figure 4

Figure 4

Figure 4. KEGG enrichment analysis of differentially expressed genes of stage 1 vs stage 2. The abscissa indicates the degree of significant enrichment.

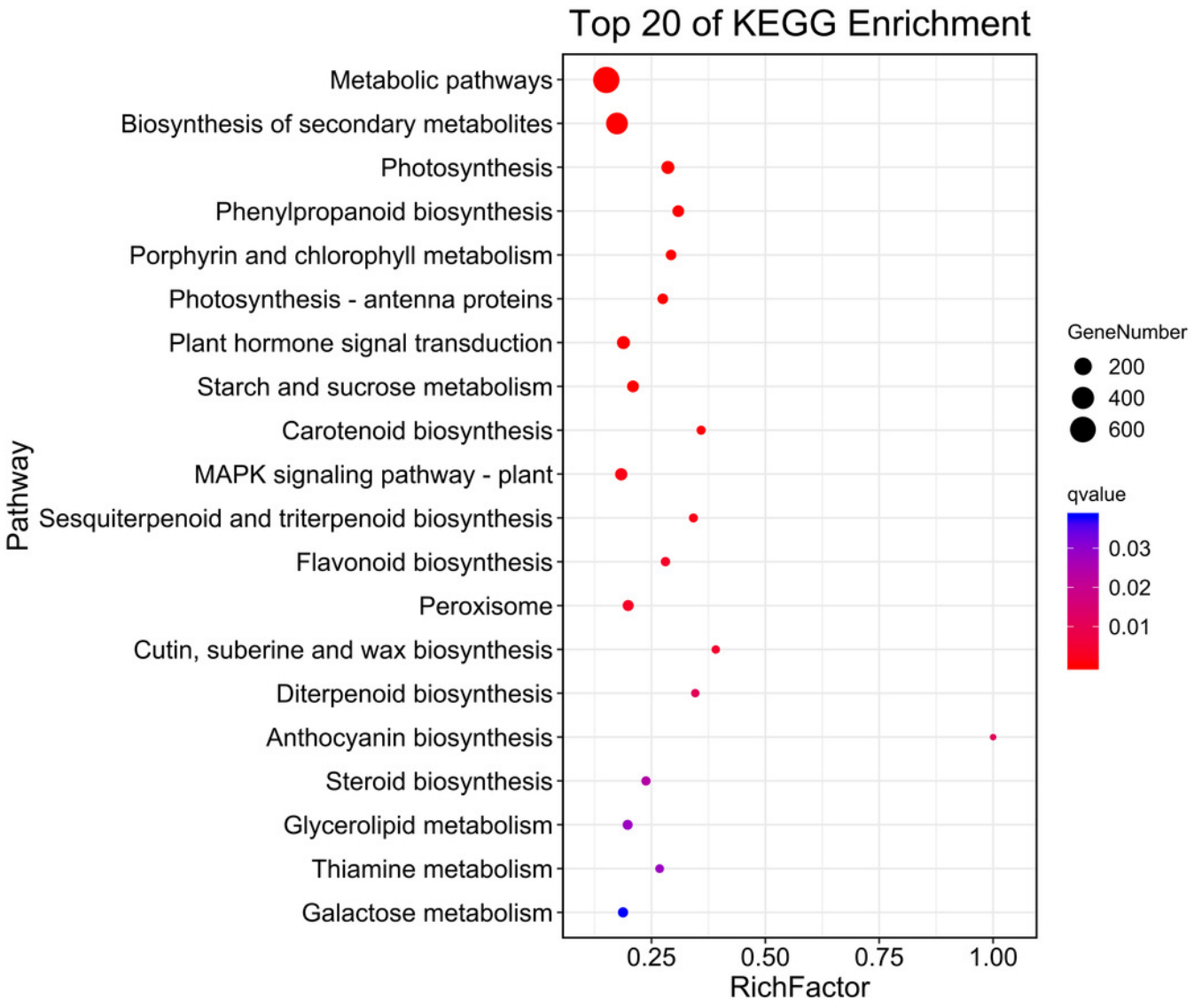


Figure 5

Figure 5

Figure 5. Analysis of leaf metabolites in different periods. (A) Differential metabolite analysis on the basis of principal component analysis (PCA) between stage 1 and stage 2. (B) OPLS-DA model plots for the comparison group stage 1 vs stage 2. (C) Differential metabolite heatmap of stage 1 vs stage 2, metabolites ID can be found in Table S2. (D) Analysis of common KEGG metabolic pathways between stage 1 and stage 2.

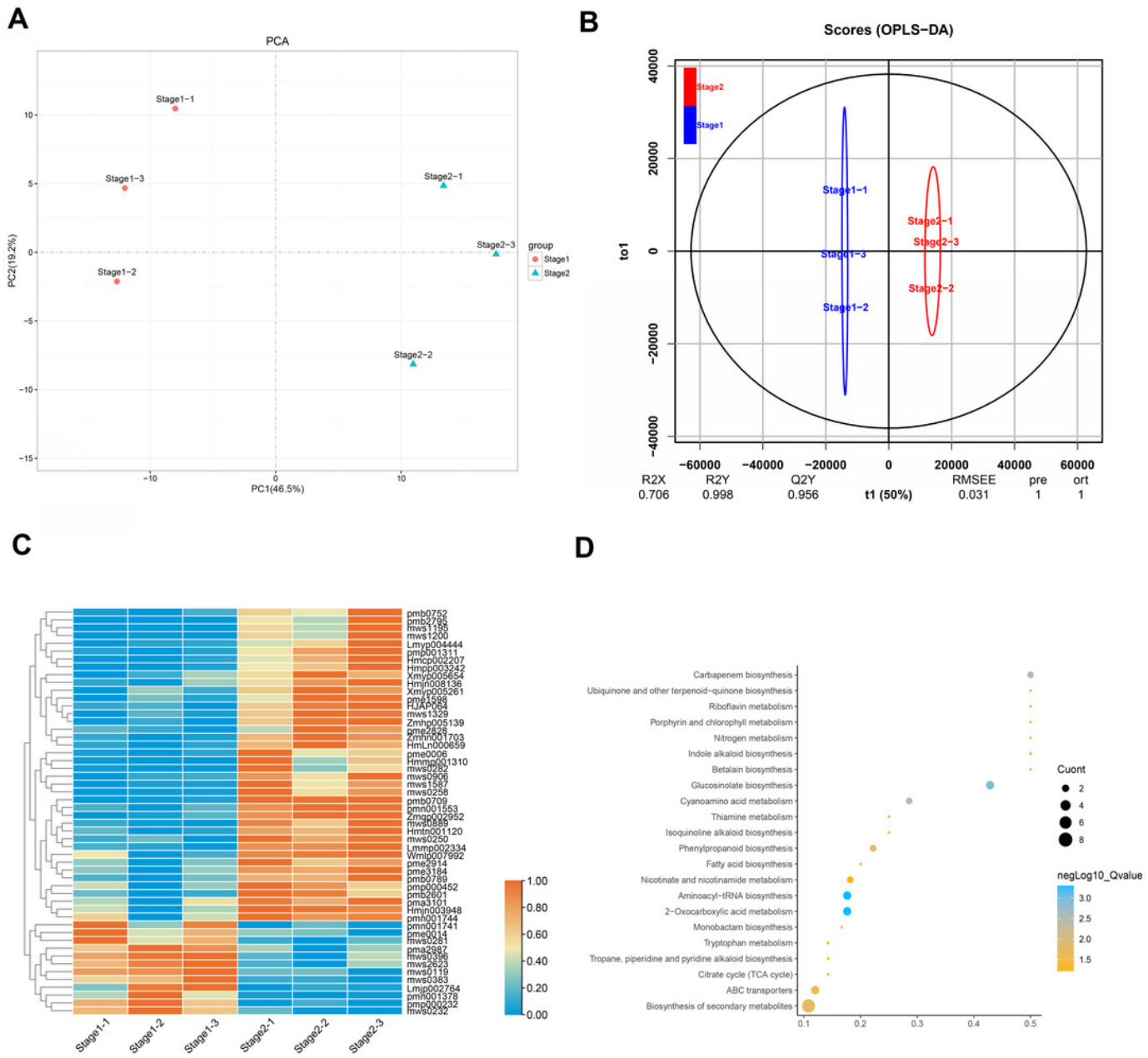

A
PCA
Stage1−1
Stage1−3
Stage2−1
Stage2−3
Stage1−2
Stage2−2
group
Stage1
Stage2
PC2(19.2%)
PC1(46.5%)
B
Scores (OPLS−DA)
Stage2
Stage1
Stage1−1
Stage2−1
Stage2−3
Stage1−3
Stage2−2
Stage1−2
to1
R2X
R2Y
Q2Y
RMSEE
pre
ort
0.706
0.998
0.956
t1 (50%)
0.031
1
1
C
pmb0752
pmb2795
mws1195
mws1200
Lmyp004444
pmp001311
Hmcp002207
Hmpp003242
Xmyp005654
Hmjn008136
Xmyp005261
pme1598
HJAP064
mws1329
Zmhp005139
pme2826
Zmhn001703
HmLn000659
pme0006
Hmmp001310
mws0282
mws0906
mws1587
mws0258
pmb0709
pmn001553
Zmqp002952
mws0889
Hmtn001120
mws0250
Lmmp002334
Wmlp007992
pme2914
pme3184
pmb0789
pmp000452
pmb2601
pma3101
Hmyp003948
pmn001744
pmn001741
pme0014
mws0281
pma2987
mws0396
mws2623
mws0119
mws0383
Lmjp002764
pmh001378
pmp000232
mws0232
1.00
0.80
0.60
0.40
0.20
0.00
Stage1-1
Stage1-2
Stage1-3
Stage2-1
Stage2-2
Stage2-3
D
Carbapenem biosynthesis
Ubiquinone and other terpenoid−quinone biosynthesis
Riboflavin metabolism
Porphyrin and chlorophyll metabolism
Nitrogen metabolism
Indole alkaloid biosynthesis
Betalain biosynthesis
Glucosinolate biosynthesis
Cyanoamino acid metabolism
Thiamine metabolism
Isoquinoline alkaloid biosynthesis
Phenylpropanoid biosynthesis
Fatty acid biosynthesis
Nicotinate and nicotinamide metabolism
Aminoacyl−tRNA biosynthesis
2−Oxocarboxylic acid metabolism
Monobactam biosynthesis
Tryptophan metabolism
Tropane, piperidine and pyridine alkaloid biosynthesis
Citrate cycle (TCA cycle)
ABC transporters
Biosynthesis of secondary metabolites
0.1
0.2
0.3
0.4
0.5
Cuont
2
4
6
8
negLog10_Qvalue
3.0
2.5
2.0
1.5

Figure 6

Figure 6

Figure 6. Association analysis of metabolomics and transcriptomics data. (A) Divergent gene and metabolite loading map of stage 1 and stage 2. (B) Network diagram of the correlation between differential gene expression and differential metabolite abundance.

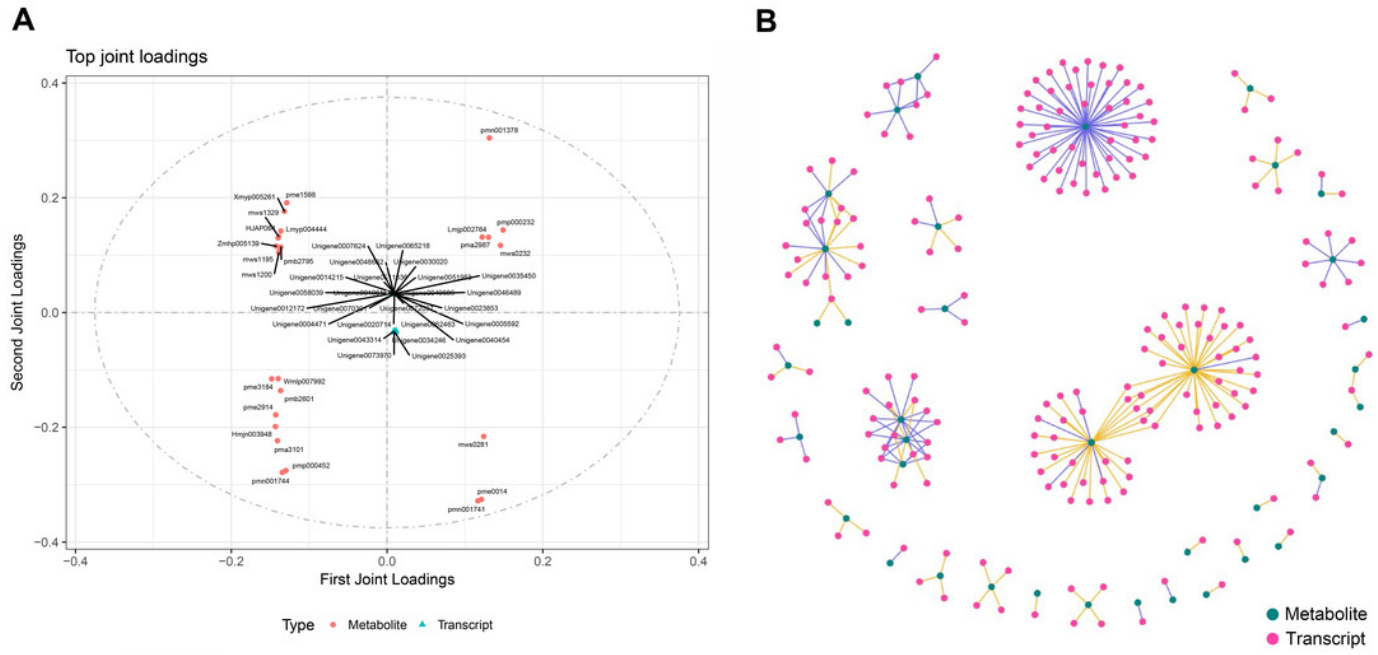
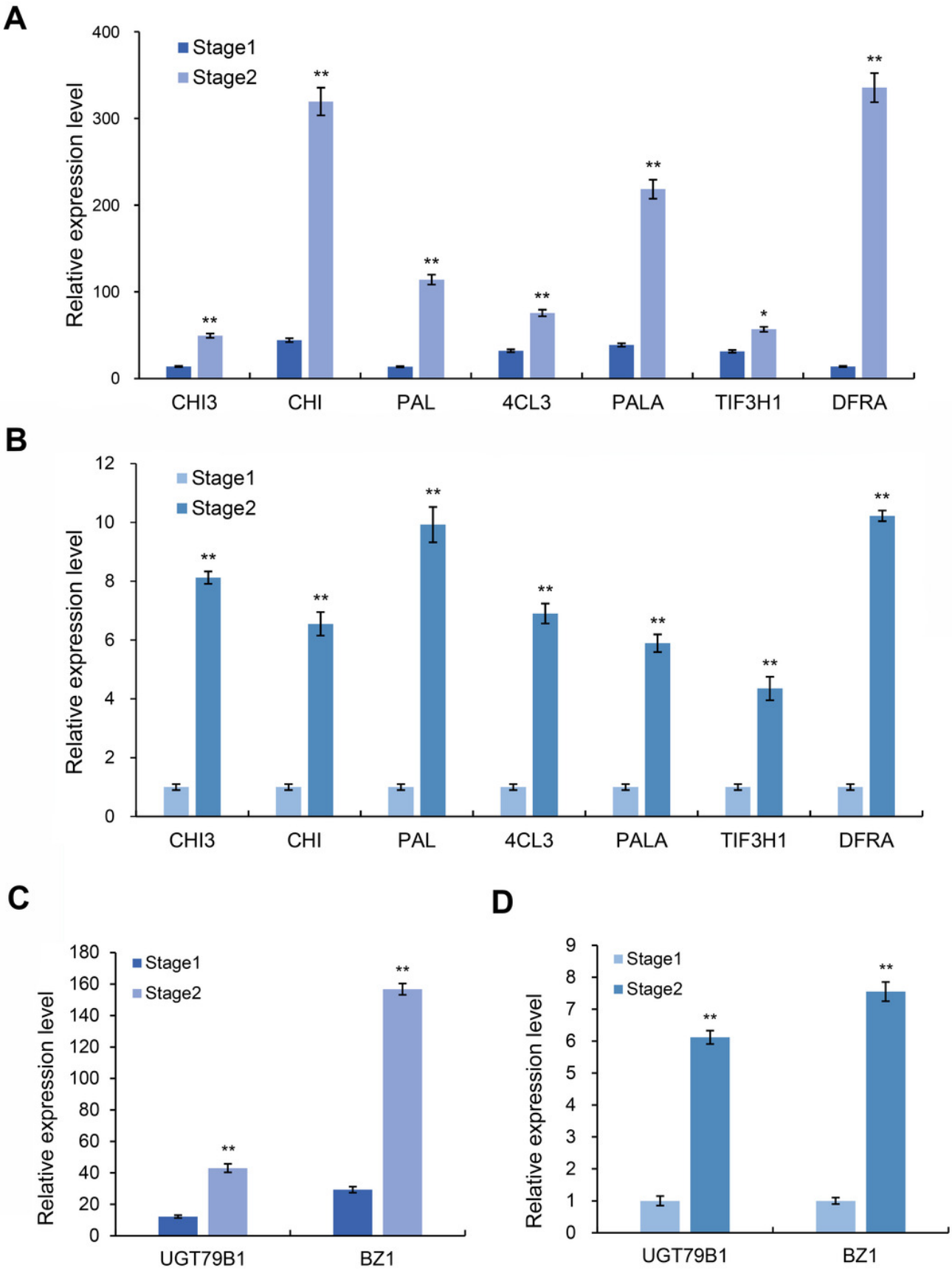


Figure 7

Figure 7

Figure 7. Expression level of anthocyanin and synthesis pathway-related genes. (A) Transcriptome data of anthocyanin synthesis pathway-related genes. (B) qRT-PCR results of anthocyanin synthesis pathway-related genes in different stages. Error bar represents the mean ± SD of three independent biological replicates. (C) Transcriptome data of anthocyanidin synthesis pathway-related genes *BZ1* and *UGT79B1*. (D) qRT-PCR results of *BZ1* and *UGT79B1* in two stages. Error bar represents the mean ± SD of three independent biological replicates.


A
Stage1
Stage2
Relative expression level
400
300
200
100
0
**
**
**
**
**
*
**
CHI3
CHI
PAL
4CL3
PALA
TIF3H1
DFRA
B
Stage1
Stage2
Relative expression level
12
10
8
6
4
2
0
**
**
**
**
**
**
**
CHI3
CHI
PAL
4CL3
PALA
TIF3H1
DFRA
C
Stage1
Stage2
Relative expression level
180
160
140
120
100
80
60
40
20
0
**
**
UGT79B1
BZ1
D
Stage1
Stage2
Relative expression level
9
8
7
6
5
4
3
2
1
0
**
**
UGT79B1
BZ1